INVESTIGATING THE STRUCTURE AND FUNCTION OF ARAC THROUGH UNCONVENTIONAL APPROACHES

by
Arjun Tambe

A thesis submitted to Johns Hopkins University in conformity with the requirements for the degree of Master of Science.

Baltimore, Maryland

April, 2017

ABSTRACT

The arabinose operon in *Escherichia coli* encodes the genes necessary to metabolize the sugar arabinose. AraC, the operon's regulatory protein, induces transcription of the *araBAD* genes in the presence of arabinose, but repress transcription through DNA looping in the absence of arabinose. The operon serves as a model system for studying the regulation of gene expression at a molecular level. However, the structure and mechanisms of AraC remain incompletely characterized. AraC exists as a relatively rigid dimerization domain (DD) connected to two DNA-binding domains (DBDs) by flexible linkers. Crystallization of full-length AraC has been repeatedly unsuccessful, so new ways to observe structure must be considered.

Here, I study the loop region between two α-helices within the DD to observe whether the loop may play a role in AraC's response to arabinose. I also create and test a fluorescence-based assay to measure AraC function *in vivo*. Finally, I develop a genetic method to probe structural aspects of AraC through the identification of complementing mutations.

Some single-amino acid substitutions can be corrected by a complementing substitution on a nearby residue. Using an AraC-encoding plasmid transformed into cells lacking AraC, loss-of-function mutations can be easily generated and isolated. These mutations can be reverted to generate *araC* genes possessing both the original loss-of-function mutation and a second-site correcting mutation in a presumably proximal residue. With knowledge of the locations of complementing residues, it will be possible to identify intra- and inter-domain interactions in AraC and predict tertiary structure. Beyond the arabinose operon, adaptations of this approach in other systems could lead to

improved models of protein folding, the synthesis of engineered peptides, and incisive methods for investigating mechanistically important domain-domain interactions.

Studies in the loop of AraC showed that introducing a cysteine into residues 144 and 145 caused aggregation and a non-functional protein. However, randomization of loop residues generated mutants that were able to both induce and repress transcription, indicating that mutation in the loop without the loss of function is possible. Lastly, I used multiple approaches to optimize a genetic method of generating complementing mutations. The system I propose lays the groundwork for large-scale genetic screens to obtain information about AraC.

PREFACE

I began working in the Schleif lab in May of 2014. At the time, I was a freshman in college who had not taken biochemistry, cell biology, or genetics. For the first year that I worked in the lab, I was admittedly conducting experiments whose principles I did not yet fully understand. Furthermore, the breadth of experiments used to study AraC meant that the moment I finally understood a concept, I was confronted with another alien one. A naïve person might say that research on a prokaryotic operon ought to be relatively straightforward – that at a time when science has progressed to CRISPR and *in vivo* gene editing, surely studying the best-characterized bacterial species should have few difficulties and surprises.

I quickly learned that this was not the case. In presenting us with AraC, Nature was kind enough to provide a protein that led to the discovery of DNA looping and demonstrated a system in which a single protein acts to actively induce and actively repress transcription. Unfortunately, Nature was not kind enough to make that protein easy to work with. I was often working more to make an experiment work than to actually obtain a result. The unexpected result that I did find, however, was how much I learned while diagnosing each problem. Although there were many experiments that did not have the outcome I had hoped for, no experiment was worthless. Even if it did not produce a finding of significance, every experiment still contributed to my growth as a scientist. As I turn to a new chapter of my life, medicine, I can say with confidence that the past four years have taught me to consider the intricate details of a biological system, not just the superficial interactions that are often described in textbooks.

ACKNOWLEDGEMENTS

Although my name appears alone on this thesis, I would like to thank many other individuals who have helped to produce this work. First, Bob Schleif, for allowing me to work in the laboratory as both an undergraduate and graduate student, and for his guidance. Bob has taken a very active role in the development of the students in the lab, including me. Second, I would like to thank Michael Rodgers, for always answering any question I threw at him, no matter when. Third, my fellow graduate students – Ory Mayberry and Matthew Brown – for giving me company and providing their input. Fourth, to the faculty of the JHU Biology Master's program: Katie Tifft and Robert Horner. Finally, thank you to my family and friends for their support over the past four years.

TABLE OF CONTENTS

Section	Page
Title Page	i
Abstract	ii
Preface	iv
Acknowledgements	v
Table of Contents	vi
List of Tables	vii
List of Figures	viii
Chapter 1: Introduction	1
Chapter 2: Mutagenesis in the Loop Region	
Introduction	6
Materials and Methods	8
Results and Discussion	14
Summary and Conclusion of Loop Experiments	21
Chapter 3: Developing a Fluorescence-Based Assay to Measure AraC Function	
Introduction	23
Materials and Methods	23
Results and Discussion	25
Chapter 4: Using Genetics to Understand Protein Structure through Complementing Mutations	
Introduction	27
Materials and Methods	28
Results and Discussion	35
Summary of Mutagenesis Screens	47
References	48
Curriculum Vitae	52

LIST OF TABLES

Table	Description	Page
Table 1	The AraC sequence in residues 143-146 and growth on minimal arabinose media after randomization of the nucleotides encoding residues 143-146 in *araC* of pBAD-GFP	18
Table 2	The RIRY mutant induces P_{BAD} to approximately 81% of WT AraC in the presence of arabinose	19
Table 3	AraC proteins with mutations in residues 143-146 retained the ability to repress transcription from P_C	21
Table 4	Typical values from the pBAD-mCherry fluorescence assay of AraC function	26
Table 5	$AraC^+$ revertants that were sequenced in mutagenesis screen 1	36
Table 6	$AraC^+$ revertants that were sequenced in mutagenesis screen 2	39
Table 7	$AraC^-$ candidates that were sequenced in mutagenesis screen 3	41
Table 8	$AraC^-$ candidates that were sequenced in mutagenesis screen 4	44

LIST OF FIGURES

Figure	Description	Page
Figure 1	An overview of AraC-mediated regulation of transcription	3
Figure 2	Crystal structure of the AraC dimerization domain when bound to arabinose	7
Figure 3	The pBAD-GFP Plasmid	8
Figure 4	Chapter 4 methods overview	35

CHAPTER 1: INTRODUCTION

The arabinose operon in *Escherichia coli* encodes the proteins necessary to metabolize L-arabinose, a five-carbon sugar. Acting together, eight genes in the operon allow the bacterium to import and catabolize the sugar, as well as regulate the levels of the associated proteins. Arabinose enters a cell through one of two methods: a low-affinity, high-capacity transporter (AraE) or a high-affinity, low-capacity transport system (AraF, AraG, and AraH) (Khlebnikov, Risa, Skaug, Carrier, & Keasling, 2000). AraE is regulated by the P_E promoter, while AraF, AraG, and AraH are regulated by a mutual P_{FGH} promoter. After arabinose uptake, three enzymes (AraB, AraA, and AraD) metabolize intracellular arabinose into D-xylulose 5-phosphate, which then enters central metabolism pathways; all three enzymes are regulated by a mutual P_{BAD} promoter (R. Schleif, 2003). Finally, the protein AraC – with the *araC* gene controlled by its own P_C promoter – regulates the operon, functioning as both a transcriptional activator and repressor. When arabinose is present, AraC binds the sugar and induces transcription of the *araB*, *araA*, and *araD* (*araBAD)* genes by activating the P_{BAD} promoter (Figure 1). In the absence of arabinose, AraC represses transcription. This dual activating-repressing ability of AraC is explained by the phenomenon of DNA looping, which results from the protein's structure and its interaction with DNA (R. Schleif, 2010).

AraC is a homodimer; each monomer is composed of a dimerization domain (DD) and a DNA-binding domain (DBD) connected by a six-residue linker. The DD portion of each monomer contains two α-helices connected by a short loop, a β-barrel that forms the arabinose-binding pocket, and an N-terminal arm (Soisson, MacDougall-Shackleton, Schleif, & Wolberger, 1997). The DBDs interact with both RNA polymerase and DNA to

regulate transcription. In the absence of arabinose, AraC forms a DNA loop by binding to two half-sites, which are segments of DNA named because each alone constitutes half of a complete AraC binding. Without arabinose, AraC binds to the $araO_2$ and $araI_1$ half-sites. These sites are 210 DNA base-pairs apart, so the binding of AraC to both sites forces a looping conformation of the DNA (Lobell & Schleif, 1990). The loop prevents RNA polymerase from accessing both the P_C and P_{BAD} promoters. Thus, in the absence of arabinose, AraC represses its own transcription in addition to repressing transcription of the *araBAD* genes (Figure 1) (R. Schleif, 2003). However, when arabinose is present, it binds to AraC and changes the conformation of the protein, so AraC binds instead to two adjacent half-sites: $araI_1$ and $araI_2$ (Wu & Schleif, 2001b). In this setting, the DNA loop is not formed, and AraC forms a complex with cyclic AMP binding protein (CAP) and RNA polymerase to induce transcription of *araBAD* and *araC* (Figure 1) (R. Schleif, 2003).

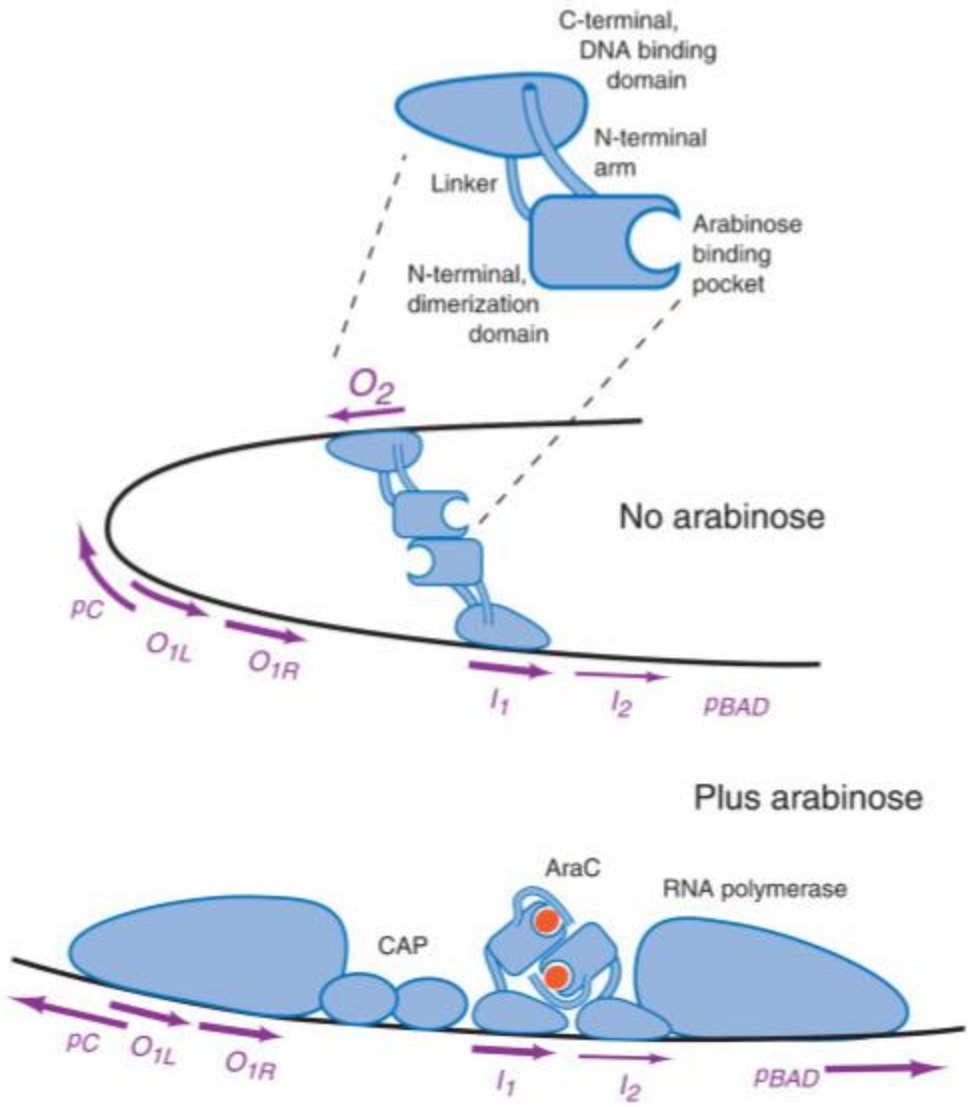

Figure 1. An overview of AraC-mediated regulation of transcription. In the absence of arabinose, AraC represses transcription by binding to the *araO*$_2$ and *araI*$_1$ half-sites, forming a DNA loop. In the presence of arabinose, AraC binds instead to the *araI*$_1$ and *araI*$_2$ half-sites, breaking the loop and inducing transcription. Arabinose is indicated as a red circle (R. Schleif, 2003).

The conformational change of AraC induced upon the binding of arabinose is not yet fully understood. Previous research has indicated that the N-terminal arms play a role in orienting the DBDs to bind half-sites. It has been thought that in the absence of arabinose, each arm adopts a structure on the DD upon which the DBDs bind, creating a rigid protein that cannot bind the adjacent *araI*$_1$ and *araI*$_2$ half-sites (R. Schleif, 2010). When AraC binds arabinose, the arms restructure and allow the DBDs to have more

flexibility. As a result, AraC is able to bind to the adjacent half-sites (Rodgers, Holder, Dirla, & Schleif, 2009). The linkers connecting each DBD to the DD are thought to play a role in the conformational change as well. The linkers are only 6 amino acids in length, but a engineered protein consisting of two AraC DBDs connected by flexible 13- or 19-residue linkers bound to the adjacent $araI_1$ and $araI_2$ half-sites. This suggests that in the absence of arabinose, the rigid orientation of the DBDs is what prevents binding to $araI_1$-$araI_2$, leaving the protein to form the repression loop by binding to $araO_2$-$araI_1$ instead (Harmer, Wu, & Schleif, 2001).

The mechanisms by which arabinose binding shifts AraC from DNA looping to binding to the adjacent $araI_1$-$araI_2$ half-sites are incompletely understood. My work has consisted of three different approaches to research regulation of the arabinose operon by AraC. All three approaches share the overarching objective of understanding AraC's structure and its function. The dual inducing-repressing role that AraC plays makes the arabinose operon in *E. coli* different from many other model systems, such as the *lac* operon. The simplicity of the arabinose operon's regulation at the macromolecular level makes it an excellent system for studying gene expression using several approaches, including biophysical, biochemical, and genetic studies. The principles demonstrated in the arabinose operon can be applied to transcriptional regulation across many species. Furthermore, if sufficient molecular understanding is reached, the allosteric principles in AraC can be used to engineer other proteins, as has already been crudely done in one instance (R. Schleif, 2003).

Despite the overall simplicity of the arabinose operon, the AraC protein has been difficult to work with biochemically. Repeated efforts over the course of multiple

decades have failed to yield crystals of full-length AraC that can be used for X-ray diffraction (R. Schleif, 2010). The dimerization domain has been successfully crystallized in the absence and presence of arabinose (Soisson et al., 1997; Weldon, Rodgers, Larkin, & Schleif, 2007). The DNA-binding domains resist crystallization, but their structure was determined by NMR (Rodgers & Schleif, 2009). AraC's regulation of transcription seems to involve interaction between the dimerization domain and DNA-binding domains, so knowledge of the DD and DBD structures isolated from each other is insufficient to form a comprehensive understanding of AraC. Thus, new approaches are required to study the protein. I have used biochemical, *in vivo*, and genetic assays not only to investigate AraC, but also to provide tools that will aid future studies – both within the arabinose operon and in other systems.

CHAPTER 2: MUTAGENESIS IN THE LOOP REGION

Introduction (Chapter 2)

Previous studies have shown that the orientation of the DNA-binding domains (DBDs) in AraC influences whether the protein induces or represses transcription from the P_{BAD} and P_C promoters (Wu & Schleif, 2001a). The orientation of the DBDs is apparently controlled by the approximately 18-residue N-terminal arms of each monomer. In the absence of arabinose, the arms are postulated to interact with the DBDs, holding them in a sufficiently rigid conformation that formation of a DNA loop is favored. Upon the binding of arabinose, the arms adopt a different conformation, freeing the DNA-binding domains such that they reorient and bind to the adjacent $araI_1$ and $araI_2$ half-sites (R. Schleif, 2010). Not surprisingly, mutation in the arms or deletion of the arms results in an inability of the protein to repress transcription. Without the arm functioning properly, AraC binds $araI_1$-$araI_2$ and induces transcription by default. As this induction does not require arabinose, the expression is called constitutive (Saviola, Seabold, & Schleif, 1998).

The crystal structure of AraC's DD in the presence of arabinose shows that the 8-residue inter-domain linker found on each monomer (residues 167-174) lies proximal to both the N-terminal arm and the loop between the two α-helices of the opposite monomer (Figure 2). Recent results indicate that the linker changes structure upon the binding of arabinose to AraC (Malaga et al., 2016).

The loop constitutes residues 143-146. Although short, the relatively unstructured loop could interact with the opposite linker and help to orient the DBD in its repressing or inducing state, perhaps as a result of the arm's change in conformation upon arabinose

binding. The goal of my study was to observe whether there is a conformational change in the loop upon the binding of arabinose by labeling a loop residue for use in fluorescence anisotropy experiments.

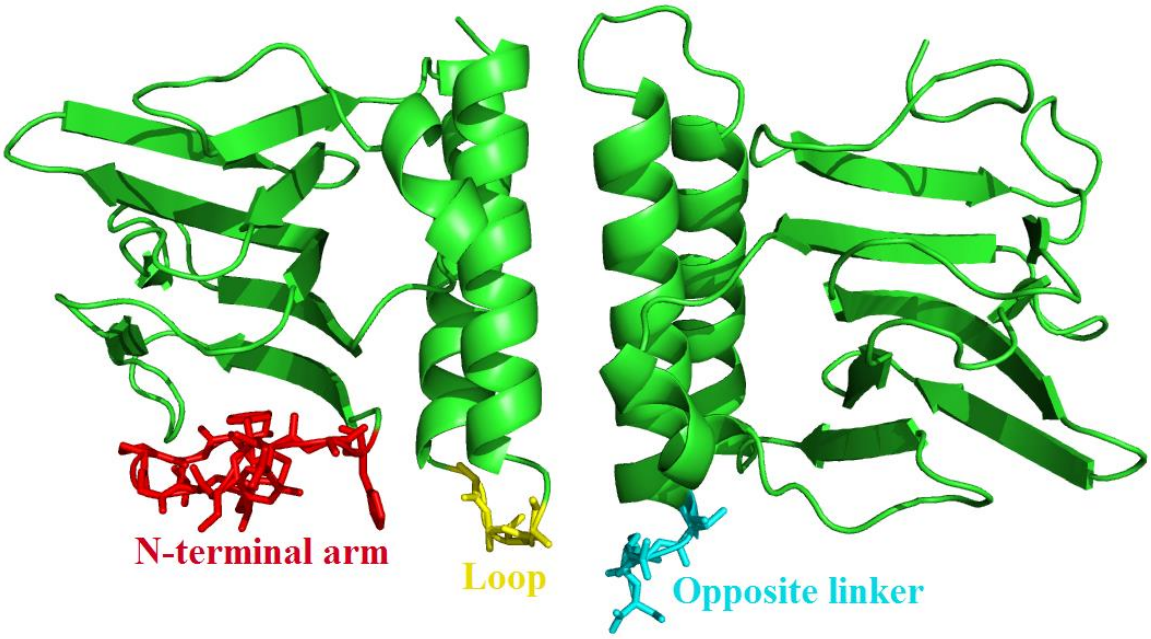

Figure 2. Crystal structure of the AraC dimerization domain when bound to arabinose. One arm (red), one loop (yellow), and one linker (cyan) are labeled. The loop lies between the arm of the same monomer and the linker of the opposite monomer (Soisson et al., 1997). A conformational change in the arm could be translated through the loop to the opposite linker, affecting the orientation of the DNA-binding domain attached to the linker.

AraC has been previously labeled with 5-({2-[(iodoacetyl)amino]ethyl}amino)-naphthalene-1-sulfonic acid (IAEDANS) on cysteine residues introduced into the interdomain linker (Malaga et al., 2016). The fluorescent label can be used in subsequent anisotropy experiments. To introduce IAEDANS into the loop, I mutated residues E144 and G145 to cysteine, providing a labeling site for the fluorophore. Then, I overexpressed and purified the mutant proteins.

The resulting proteins did not retain functionality, implying that the loop cannot tolerate cysteine residues. Randomizing residues 143-146 yielded non-functional AraC mutants and mutants that could both induce and repress transcription *in vivo*. These

findings suggest that the exact amino acid sequence of the loop may not be important for AraC function, but not all mutations are possible. The function of the loop in the AraC response to arabinose seems to be small, if it plays any role at all.

Materials and Methods (Chapter 2)

Plasmids

The pBAD-GFP (*amp*^{r}) plasmid was previously described as pBAD-AraC-GFP (Timmes, Rodgers, & Schleif, 2004). The plasmid was briefly sold by Clontech® (Mountain View, CA). It contains the *GFP* gene driven by the P_{BAD} promoter, the *araC* gene driven by the P_C promoter, and the ampicillin resistance gene β-lactamase (Figure 3).

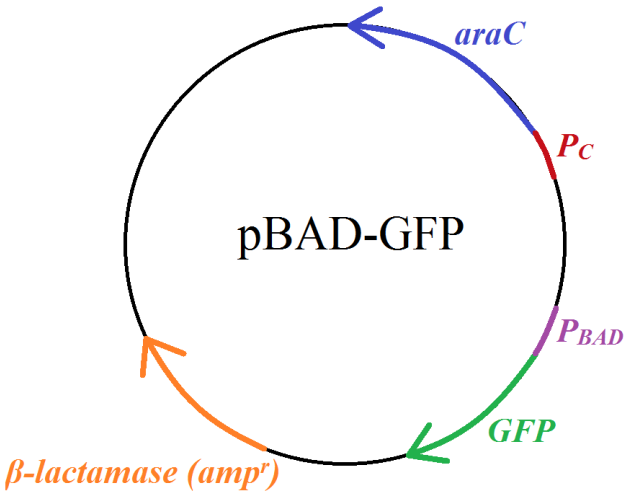

Figure 3. The pBAD-GFP plasmid.

The pET21-DD (*amp*^{r}) plasmid was previously created, referred to as AraCTF (Weldon et al., 2007). The plasmid allows for the overproduction of the dimerization domain of AraC.

The pET24d (*kan*^{r}) plasmid was previously created for the overexpression of full-length AraC protein (Rodgers & Schleif, 2009).

The pWR03 (*amp*^{r}) plasmid was previously created by cloning a DNA fragment containing the entire AraC coding region (nucleotides 1-873) into the *Nco*I and *Sac*I sites of pSE380, obtained from Invitrogen (San Diego, CA).

Strains

Competent DH5α cells were obtained from Invitrogen (San Diego, CA).

BL21(DE3) cells were used to overexpress AraC protein with the pET21-DD and pET24d plasmids (Studier, Rosenberg, Dunn, & Dubendorff, 1990). Competent BL21(DE3) cells were originally obtained from Invitrogen (San Diego, CA).

SH321 (*F*$^{-}$*, ΔaraC-leu1022 Δlac74, galK*$^{-}$*, str*^{r}) cells lack a chromosomal copy of AraC, as described previously (Hahn, Dunn, & Schleif, 1984).

SH10 (F^{+}, *araC:lacZ fusion, leu*$^{-}$*str*^{r}*, Δlac araD*$^{-}$*)* cells contain the *lacZ* gene instead of *araC*. The cells were previously documented, referred to as DMH90 (Hahn & Schleif, 1983).

Media

Yeast tryptone (YT) was used for rich medium as previously described, both for plates and liquid cultures. Plates were made with 15 grams of agar per liter. (R. F. Schleif & Wensink, 1981). To screen cells using a single sugar source, minimal media were used: either A+B or M10, as previously described (R. F. Schleif & Wensink, 1981). Minimal media were supplemented with leucine (50 μg/mL), vitamin B1 (10 μg/mL),

$MgSO_4$ (1 mM), and either arabinose or glycerol (2 mg/mL). IPTG was used at a concentration of 1 mM unless otherwise indicated. Antibiotics were used at the following concentrations unless otherwise indicated: ampicillin 100 μg/mL (amp), kanamycin 40 μg/mL (kan).

Cell Growth

Cells were grown at 37 ˚C unless otherwise indicated. For liquid media, cells were grown in a shaker to approximately stationary phase unless otherwise indicated. Typical growth duration was 1 day on YT medium and 2-4 days on minimal media.

Site-Directed Mutagenesis

The Agilent Technologies QuikChange™ protocol for oligonucleotide-directed mutagenesis was followed for all reactions. After thermocycling, the reaction mixture was incubated with *Dpn*I (1 μL, 37 ˚C, 2 hours) obtained from New England Biolabs® (Ipswich, MA) to digest any parental (methylated) DNA. Primers were ordered from Integrated DNA Technologies (Coralville, IA).

Transformation

Plasmid DNA (1-2 μL) was mixed with 50 μL competent *E. coli* cells (various strains) and incubated on ice for 30 min. Cells were subjected to heat-shock (42 ˚C, 20-45 seconds) and promptly returned to ice (2 min). YT medium (500 μL) was added, and cells were allowed to recover without antibiotics (37 ˚C, approx. 1 hour). Following

recovery, cells were either sterilely spread onto appropriate plates or used to inoculate liquid media (YT unless otherwise indicated).

Plasmid DNA Extraction

Plasmid DNA was extracted from cells using the Promega Wizard® *Plus* SV Minipreps DNA Purification System: Cells were used to inoculate 5 mL liquid YT media supplemented with appropriate antibiotics and other nutrients if necessary. Tubes were incubated until the cells reached sufficient growth for sequencing (37 ˚C, approx. 18 hours). Cells were spun down (3000 RPM, 10 min), and plasmid DNA was extracted following the protocol. DNA concentration was determined using a NanoDrop 2000 UV-Vis Spectrophotometer.

Sequencing

Plasmid DNA (approx. 650 ng) was sent to GENEWIZ® (South Plainfield, NJ) for sequencing with the appropriate primer.

pET21-DD Protein Overexpression and Purification

Mutant protein was overexpressed and purified using the pET21-DD vector essentially as described (Weldon et al., 2007). The E144C and G145C mutations were first introduced into the *araC* gene of pET21-DD. The primer sequence used to generate E144C was 5'-GCCGGGCAAGGGTGCGGGCGCTATTCGG-3' and its complement. The primer sequence used to generate G145C was 5'-CCGGGCAAGGAGAATGCCGCTATTCGGAG-3' and its complement. Mutated pET21-DD was transformed into competent BL21(DE3)

cells and plated onto YT/amp. A starter culture was generated using 5 mL YT/amp inoculated with a single colony from the plate, and cells were grown overnight. Two 500 mL YT/amp flasks were inoculated using the starter culture and grown to A_{550} = 0.8-1.0. IPTG (0.4 mM) was added to induce AraC overexpression, and growth continued for an additional 5 hours.

Cells with overexpressed protein were harvested via centrifugation and resuspended in lysis buffer (0.1 M NaCl, 15 mM Tris-HCl pH 8, 10 mM $MgCl_2$, 5 mM arabinose, 5% glycerol, 10 µg/mL DNase, 10 µg/mL RNase). Cells were lysed, then centrifuged (15000 RPM, 15 min). The supernatant was passed through a Ni^{2+}-NTA affinity chromatography column equilibrated with wash buffer (15 mM Tris-HCl pH 8, 100 mM NaCl, 5 mM arabinose, 10 mM imidazole). The flow-through was collected and passed through the column a second time. The column was washed with wash buffer until eluate A_{280} was below 0.05. AraC was eluted using elution buffer (15 mM Tris-HCl pH 8, 10 mM NaCl, 5 mM arabinose, 1 M imidazole).

The eluate was incubated with 1 µg trypsin per mg of estimated protein (4 ˚C, 16 hours), then soybean trypsin inhibitor was added in equimolar concentration to inhibit the digest. Final protein purification was achieved using high-performance liquid chromatography. Protein was bound to the column without salt (15 mM Tris-HCl pH 8, 0.1 mM NaN_3, 5 mM arabinose) and eluted with salt concentration increasing to 1 M NaCl.

GFP Fluorescence Assay to Measure P_{BAD} Induction

Three different plasmids were transformed into the SH321 cell line: pBAD-GFP WT, pBAD-*gfp*$^-$ (WT AraC sequence but non-functional GFP), and pBAD-GFP with the mutation 143-146RIRY ("RIRY"). Cells were grown overnight in 5 mL of liquid YT/amp media, both with and without arabinose. A fluorescence-based AraC induction assay was performed essentially as previously described (Timmes et al., 2004). Following growth, OD was calculated by measuring Abs_{600}. Cells were lysed with detergent and excited at 395 nm; GFP emission was measured at 508 nm, and cell density was measured proportional to emission at 395 nm. One 10%-transmittance neutral-density filter was used to measure cell density at 395 nm.

Calculations were performed as follows: The values from a buffer blank were subtracted from values for cell samples. The ratio of $[508nm_{sample}]/[395nm_{sample}]$ was determined. Normalized fluorescence was calculated relative to the pBAD-GFP WT Ara+ sample using the following formula: $[Ratio_{sample}]/[Ratio_{wt\text{-}Ara+}]$.

P_C Repression Assay

A repression assay was performed based on the protocol of Zhang and Bremer (1995), with solutions made as described. The assay uses *ortho*-nitrophenyl-β-galactoside (ONPG) to detect β-galactosidase activity through formation of the yellow product *ortho*-Nitrophenol. If AraC (encoded on the pBAD-GFP plasmid for this assay) maintains repressing ability, then SH10 cells (containing the P_C-*lacZ* fusion) should form little colored product in the absence of arabinose due to the lack of *lacZ* transcription. The amount of *ortho*-nitrophenol produced was quantified by absorbance at 420 nm.

Beta-galactosidase levels were measured as described, with some modifications (R. F. Schleif & Wensink, 1981). SH10 cells (with appropriate plasmids if necessary) were grown to stationary phase in YT medium; Abs_{600} was recorded. Cells (20 µL) were added to permeabilization solution (80 µL) and incubated at 30 ˚C for 25 min. Cells remained at 30 ˚C, and substrate solution (600 µL) was added. After sufficient yellow color was observed, stop solution (700 µL) was added, and samples were kept at room temperature. The time between adding substrate solution and stop solution was noted. Samples were spun in a microcentrifuge (maximum speed, 10 min). The Abs_{420} of the supernatant was recorded.

Miller Units were calculated using the formula: $M.U. = [1000*Abs_{420}] / [t*v*Abs_{600}]$, where t represents the reaction time (mins) for each sample, and v is the volume of cells used in milliliters (for this study, 20 µL). Higher Miller Units indicates an inability to repress transcription from P_C.

Results and Discussion (Chapter 2)

The E144C Mutation in AraC DD Produced Aggregating Protein

To label the loop with IAEDANS, the mutation E144C was introduced to the *araC* gene of the pET21-DD vector, which encodes the DD of AraC and provides ampicillin resistance. Mutations were introduced following the QuikChange™ protocol. Plasmids containing the mutated DNA were transformed into competent DH5α *E. coli* cells and plated on YT/amp. Isolated colonies were grown in liquid YT/amp, and their plasmid DNA was purified and sent for sequencing to verify successful mutagenesis. Following verification, plasmid DNA was transformed into competent BL21(DE3) cells

and plated onto YT/amp. Cells were grown in liquid culture, induced to overexpress the dimerization domain, and lysed.

While attempting to purify the E144C AraC protein from the lysate with affinity chromatography, significant elution was not achieved. The introduction of the cysteine at residue 144 likely disrupted protein folding, causing aggregation. Alternatively, an exposed cysteine could have still resulted in aggregated protein even if folding was not affected due to the formation of disulfide bonds. Dithiothreitol (DTT; 10 mM) was introduced to maintain a reducing environment and prevent unwanted disulfide bonds from forming, but sufficient protein elution was still not achieved. This finding suggests that introducing a cysteine at residue 144 may affect AraC structure too much for further study.

The E144C Mutation in Full-Length AraC Produced Non-Functional Protein *in vivo*

To investigate the effects of the E144C mutation *in vivo*, E144C was introduced into the pET24d plasmid, which encodes full-length AraC and provides kanamycin resistance. Mutagenesis was performed using the standard QuikChange™ protocol, and isolated colonies were selected for sequencing. Verified DNA was transformed into competent SH321 cells, which lack a chromosomal copy of *araC*. Colonies were dual-spotted onto minimal nutrient plates containing either glycerol (A+B/gly/kan) or arabinose (A+B/ara/kan) as their sole carbon source. All cells would be expected to grow on glycerol, but only transformants with functional AraC proteins should grow on arabinose because AraC is necessary to induce the expression of *araBAD*. All spotted

colonies containing the E144C mutation in *araC* grew on glycerol but not arabinose, confirming that the E144C mutation results in a non-functional AraC protein.

The G145C Mutation in AraC DD Also Produced Aggregating Protein

Because purification of the E144C mutant was unsuccessful, I instead aimed to produce the G145C mutant in pET21-DD for use in subsequent characterization experiments. No elution of protein was achieved from Ni^{2+} affinity chromatography. The purification was attempted with the reducing agents DTT (10 mM) and Tris(2-carboxyethyl)phosphine (TCEP; 1 mM) to prevent unwanted disulfide bonds, but without success. Increasing the concentration of imidazole from 1 M to 3 M also did not result in elution of the protein.

Analysis with SDS-PAGE demonstrated that protein overexpression had been achieved, but only 30% of the protein remained in the supernatant after centrifugation; the remaining 70% was aggregated in the discarded pellet. No elution was obtained from the Ni^{2+} column because most of the protein remained bound to the column beads. The insolubility of both the E144C and G145C mutants in AraC DD, in addition to the loss of AraC function in pET24d E144C, suggests that it may not be possible to introduce cysteine mutations into the loop for subsequent characterization experiments.

Randomizing Residues 143-146 Demonstrated that the Loop Can Be Mutated without Compromising Function

To investigate whether any mutations in the AraC loop region are tolerated without loss of function, the codons encoding residues 143-146 were randomized in *araC*

on the pWR03 plasmid, which contains full-length *araC* and ampicillin resistance. Randomization was achieved using QuikChange™ and the following primers: 5'-CATTAACGCCGGGCAANNNNNNNNNNNNTATTCGGAGCTGCTG-3' and its complement. Mutated DNA was transformed into the *araC*$^{-}$ SH321 cell line. Randomized colonies were dual-spotted onto A+B/gly/amp and A+B/ara/amp.

All colonies grew on A+B/gly/amp. Some, but not all, randomized colonies grew on A+B/ara/amp, indicating that mutations in the loop can still yield functional AraC protein. One colony was used for plasmid DNA extraction and sequencing of the *araC* gene; the amino acid sequence of residues 143-146 was PIDS, while the wild-type sequence is GEGR. The sequenced plasmid also had an early STOP codon in place of residue 281, in addition to the PIDS mutation. The early STOP281 codon did not affect protein function; this is consistent with previous studies, which found that some premature truncations late in the AraC primary sequence show no abnormal phenotype (Eustance & Schleif, 1996). The ability of the PIDS mutant to grow on A+B/ara/amp demonstrates that mutating the loop region of *araC* can still result in a functional protein. To increase support of this finding, a similar randomization experiment was carried out in another *araC*-encoding plasmid, pBAD-GFP.

The pBAD-GFP plasmid contains the entire *ara* regulatory region with *GFP* under control of the P_{BAD} promoter (in place of *araBAD*) and *araC* under control of the P_C promoter. The plasmid confers ampicillin resistance. Randomization of residues 143-146 was repeated in this plasmid as described for pWR03. If grown on arabinose, cells containing pBAD-GFP with a functional *araC* gene should display growth and GFP fluorescence.

The loop residues (143-146) in the *araC* gene of pBAD-GFP were mutated similarly to the randomization in pWR03. The primers used for the randomization were 5'-GGCAAATCATTAACGCCGGGCAANNNNNNNNNNNNTAATCGGAGCTGCTGGCGATAAATC-3' and its complement. Mutated plasmids were transformed into SH321 cells, and resulting colonies were dual spot-tested for growth on A+B/gly/amp and A+B/ara/amp. All colonies, including wild type, had comparable growth on A+B/gly/amp. Some, but not all, mutant colonies grew on A+B/ara/amp. Select colonies that grew on A+B/ara/amp had their pBAD-GFP plasmids sequenced (Table 1).

Table 1. The AraC sequence in residues 143-146 and growth on minimal arabinose media after randomization of the nucleotides encoding residues 143-146 in *araC* of pBAD-GFP.

Sequence in AraC Residues 143-146	Relative Growth on A+B/ara/amp
GEGR (Wild Type)	Full
TVKK	Full
RIRY	Full
TSGF	Full
PVSG	Full
TATP	Full
GARP	Full

"Full" growth indicates that growth was comparable to the growth of wild type cells on A+B/gly/amp.

The findings in pBAD-GFP support the results from using the pWR03 plasmid. Some mutations in the loop were found to eliminate AraC functionality *in vivo*. No cysteine-containing mutants with growth on minimal arabinose media were identified, suggesting that the presence of cysteine in the loop may cause the loss of AraC function. However, the growth of certain mutant colonies on minimal arabinose demonstrated that the loop region of AraC can tolerate some mutations without compromising protein function.

The 143-146RIRY Loop Mutant Had 81% of WT P_{BAD} Induction

Of the mutants found in the previous screen, pBAD-GFP with RIRY substituted for residues 143-146 ("RIRY") was chosen for further study. RIRY was selected as a representative plasmid that encoded a functional *araC* despite mutation in the loop residues.

Three different plasmids were transformed into the SH321 cell line: pBAD-GFP WT, pBAD-*gfp*$^-$ (WT AraC sequence but non-functional GFP), and RIRY. Following growth, the GFP fluorescence assay was performed (Table 2).

Table 2. The RIRY mutant induces P_{BAD} to approximately 81% of WT AraC in the presence of arabinose.

Sample	395 nm	508 nm	Ratio	OD	395 nm/OD	Norm. Fluorescence
WT Ara+	386181	1061720	2.75	9.7	39649	1.00
WT Ara-	439605	137212	0.31	7.6	57919	0.11
gfp$^-$ Ara+	359087	136715	0.38	8.2	43685	0.14
gfp$^-$ Ara-	432063	138105	0.32	7.0	61989	0.12
RIRY Ara+	427838	947120	2.21	9.0	47538	0.81
RIRY Ara-	434285	150310	0.35	7.4	58450	0.13

Values under 395 nm and 508 nm represent single-photon counts. 395 nm measures cell density, while 508 nm measures GFP fluorescence. "Ratio" was calculated using the following formula: $[508nm_{sample}]/[395nm_{sample}]$. OD is the absorbance of the cells at 600nm before lysis. Normalized fluorescence ("Norm. Fluorescence") was calculated using the following formula: $[Ratio_{sample}]/[Ratio_{wt\text{-}Ara+}]$.

The RIRY mutant, in the presence of arabinose, induced GFP fluorescence approximately 81% of the WT in arabinose. The background fluorescence, shown by the pBAD-*gfp*$^-$ samples, was approximately 12-14%. The ratio of photon counts at 575 nm to OD (Abs_{600}) should be approximately equal within the same sample of cells. However, lysing the cells may have affected the emission. The approximately equivalent calculated normalized fluorescence values for uninduced wild type cells (WT Ara-) and *gfp*$^-$ cells demonstrates a drawback of the GFP fluorescence assay. Emission profiles for GFP

overlap with cell autofluorescence, making the assay lack precision that would be required for thorough characterization of mutant AraC function. This problem will be addressed in Chapter 2 of the thesis. Despite the imprecision, this assay still demonstrates that the RIRY mutant retains significant P_{BAD} induction ability. Future studies could repeat the GFP assay to observe the induction achieved with other AraC loop mutants.

Mutants of AraC in Residues 143-146 Maintained the Ability to Repress Expression

To observe the ability of loop mutants to repress transcription, pBAD-GFP WT, pBAD-*gfp*$^-$, and pBAD-GFP plasmids with the following mutations of AraC residues 143-146 were transformed into the SH10 cell line: RIRY, TSGF, TATP, and GARP. Transformants were grown in 5 mL of liquid YT/amp overnight. A sample of plain SH10 cells (No Vector) was also grown in 5 mL YT, but without ampicillin. Cells were used in a repression assay at 30˚C, which has been previously performed with AraC (Reed & Schleif, 1999).

The repression assay was conducted using two samples of each cell type. Higher Miller Units indicates an inability to repress transcription. The results of the repression assay are summarized in Table 3.

Table 3. AraC proteins with mutations in residues 143-146 retained the ability to repress transcription from P_C.

Sample	Abs_{600}	Time (min)	Abs_{420}	Miller Units
No Vector 1	7.12	33.2	1.34	283.78
No Vector 2	7.12	33.2	1.46	308.13
GFP WT 1	5.22	63.5	0.14	21.12
GFP WT 2	5.22	63.5	0.15	22.63
gfp⁻ 1	6.00	63.5	0.13	16.80
gfp⁻ 2	6.00	63.5	0.13	16.67
RIRY 1	6.40	63.5	0.26	32.36
RIRY 2	6.40	63.5	0.26	31.50
TSGF 1	6.56	63.5	0.16	19.69
TSGF 2	6.56	63.5	0.17	20.41
TATP 1	5.92	63.5	0.09	11.31
TATP 2	5.92	63.5	0.08	11.04
GARP 1	6.56	63.5	0.10	11.76
GARP 2	6.56	63.5	0.10	11.52

Miller Units were calculated using the formula: M.U. = $[1000*Abs_{420}] / [t*v*Abs_{600}]$, where t represents the reaction time (mins) for each sample, and v is the volume of cells (20 µL). Higher Miller Units indicates an inability to repress transcription.

RIRY was the only mutant with higher Miller Units than the GFP WT control, and the difference between the values – approximately 32 for RIRY and 22 for WT – was marginal. All other mutants displayed repression ability comparable to or better than the WT. These data indicate that AraC with randomizations within the loop region retains the ability to repress P_{BAD} through DNA looping.

Summary and Conclusion of Loop Experiments (Chapter 2)

Neither the E144C nor the G145C mutant in pET21-DD could be purified, likely due to misfolding, thus precluding further study such as fluorescence anisotropy experiments. Consistent with this result, the E144C mutant of AraC in pET24d transformed into SH321 cells exhibited no growth on minimal arabinose media, suggesting that an introduced cysteine in the loop could affect protein structure and/or function. Nevertheless, randomization of residues 143-146 demonstrated that

functionality is preserved when the residues are mutated and AraC function is assessed *in vivo*. Furthermore, activity based on GFP fluorescence indicated that the RIRY mutant can induce P_{BAD} transcription to approximately 81% of the level of WT. A repression assay of P_C demonstrated that all four mutants of the 143-146 residues maintained the ability to repress transcription.

These findings suggest that the loop may not play a significant role in the conformational change of AraC upon binding of arabinose, or that the role of the loop is not significantly sensitive to its sequence. Nevertheless, the introduction of cysteine into the loop seems to significantly affect AraC structure and function. Few labeling options other than cysteine labeling exist, so attempting to introduce a cysteine into the loop to observe how the loop may change upon the binding of AraC to arabinose may still be worthwhile. Modifying the purification method or introducing a cysteine in residues 143/146 could allow for successful labeling.

CHAPTER 3: DEVELOPING A FLUORESCENCE-BASED ASSAY TO MEASURE ARAC FUNCTION

Introduction (Chapter 3)

In the presence of arabinose, AraC induces the expression of the *araBAD* genes, which encode the enzymes responsible for metabolizing arabinose. *araA* encodes L-arabinose isomerase, which catalyzes the conversion of L-arabinose to L-ribulose. The arabinose isomerase assay has been used to measure the activity of AraC *in vivo*, as previously described (R. F. Schleif & Wensink, 1981). However, the assay is subject to substantial errors and possesses a narrow range of linearity. The pBAD-GFP construct contains *araC* under control of the P_C promoter and GFP under control of the P_{BAD} promoter. pBAD-GFP fluorescence assays have been used to measure AraC activity, as in Chapter 1. However, the GFP fluorescence assay is not sufficiently sensitive for quantitation of P_{BAD} activities below approximately 25% of full induction; this is due to a high fluorescence background of cellular components in the same wavelength range as emission from GFP. Therefore, I developed an assay to measure the expression off the P_{BAD} promoter through the red fluorescent protein, mCherry, whose emission is in a range with much less interference from cellular components.

Materials and Methods (Chapter 3)

Preparation of *mCherry*

The plasmid pME-mCherry was obtained from the lab of Marnie Halpern (Carnegie Institute of Embryology, Baltimore, MD). The *mCherry* gene was amplified from pME-mCherry using PCR (Agilent Technologies, Santa Clara, CA), simultaneously introducing a *Nhe*I restriction site and an *Eco*RI restriction site. The primers used were 5'-

GGCTCCACCATGGCTAGCAAGGGCGAGG-3' and 5'-

CCGAGCTCGAATTCATTACTTGTACAGC-3'.

To prepare the *mCherry* gene for cloning into the pBAD-mCherry plasmid, the PCR product was triple-digested with the restriction enzymes *Nhe*I, *Eco*RI, and *Dpn*I, all obtained from New England Biolabs® (Ipswich, MA), following the protocols indicated. *Dpn*I was used to digest parental DNA.

Preparation of the pBAD-GFP Plasmid

The pBAD-GFP plasmid was double-digested with the restriction enzymes *Nhe*I and *EcoR*I, obtained from New England Biolabs® (Ipswich, MA), following the protocols indicated. The product was treated with thermosensitive alkaline phosphatase (TSAP; Promega, Madison, WI), following the protocols indicated, to prevent regeneration of the parental plasmid in the subsequent ligation step.

Ligation

The digested *mCherry* product and the TSAP-treated pBAD-GFP plasmid were ligated following the Quick Ligation™ protocol from New England Biolabs® (Ipswich, MA).

Final QuikChange™ to Restore *mCherry*

A final QuikChange™ reaction (Agilent Technologies, Santa Clara, CA) was performed to restore the original coding sequence of the *mCherry* gene, which was earlier mutated while introducing the *Nhe*I site. The resulting plasmid was pBAD-mCherry. The

primers used for the final QuikChange™ correction reaction were 5'-GGAGATATACATATGGTGAGCAAGGGCGAGG-3' and its complement.

mCherry Fluorescence-Based Assay of AraC Function

pBAD-mCherry was transformed into competent SH321 cells and plated onto YT/amp. A single colony was used to inoculate one 5 mL tube of liquid YT/amp ("uninduced") and one 5 mL tube of liquid YT/amp supplemented with arabinose ("induced"). SH321 cells without any plasmid ("Plain SH321") were grown in 5 mL liquid YT. Cells were grown to stationary phase, then placed on ice. Measurements were taken using a fluorimeter (50 µL cells in 3 mL M10 buffer) (R. F. Schleif & Wensink, 1981) with two detectors. Cells were excited with 575 nm light. Detector A measured emission at 575 nm to obtain cell density, and detector B measured emission at 605 nm to observe mCherry fluorescence. Two 10%-transmittance neutral-density filters were placed on detector A. No polarizers were used. A buffer blank was also measured.

Calculations were performed as follows: values from the buffer blank were subtracted from all sample measurements. The ratio of $[605nm_{sample}]/[575nm_{sample}]$ was determined. The "ratio" value for Plain SH321 cells was subtracted from the "ratio" value for uninduced and induced cells ("Minus SH321"). Finally, induction ratio ("induction") was calculated using the following formula: $[\text{Minus SH321}_{induced}]/[\text{Minus SH321}_{uninduced}]$.

Results and Discussion (Chapter 3)

An induction ratio of approximately 50 was achieved; typical values found for cells in the assay, along with subsequent calculations, are listed in Table 4.

Table 4. Typical values from the pBAD-mCherry fluorescence assay of AraC function.

Sample	575 nm	605 nm	Ratio	Minus SH321	Induction
Plain SH321	225000	12000	0.053		
Uninduced	180000	25000	0.139	0.086	
Induced	125000	550000	4.4	4.347	50.54

Values under 575 nm and 605 nm represent single-photon counts. 575 nm measures cell density, while 605 nm indicates mCherry fluorescence. An induction ratio of approximately 50 was able to be achieved. "Ratio" corresponds to [$605nm_{sample}$]/[$575nm_{sample}$]. "Minus SH321" corresponds to [$Ratio_{sample} - Ratio_{Plain\ SH321}$]. Induction corresponds to [Minus $SH321_{induced}$]/[Minus $SH321_{uninduced}$].

The creation of the pBAD-mCherry plasmid and the development of the fluorescence-based assay provides an alternative to arabinose isomerase assays. The fluorescence assay provides a quick method to assess AraC function, indicated by transcription from P_{BAD} and resulting mCherry fluorescence when cells are grown with arabinose. Previously, GFP fluorescence has been used, as documented in Chapter 1 of this thesis and previous studies (Timmes et al., 2004). Substitution of GFP for mCherry allows for a more precise assay, minimizing the magnitude of unwanted cell autofluorescence. Future studies should refine this method further by comparing induction ratios of well-characterized AraC mutants (on the pBAD-mCherry plasmid, in SH321 cells) with the induction ratio of wild type pBAD-mCherry plasmid in SH321 cells.

CHAPTER 4: USING GENETICS TO UNDERSTAND PROTEIN STRUCTURE THROUGH COMPLEMENTING MUTATIONS

Introduction (Chapter 4)

The structure and function of proteins are intricately related. Simply understanding a protein's structure can sometimes shed light on how the protein acts mechanistically. In the case of AraC, inter-domain interactions are thought to dictate the protein's behavior. However, due to AraC's resistance to crystallization, these structural features cannot be detected through classical methods for structure determination. Here, I demonstrate an approach that, when refined, can use simple genetics to gain structural information about AraC. Additionally, this approach should be readily adaptable to other proteins.

Single amino acid substitutions in proteins have the potential to reduce protein function. Introducing a different amino acid alters local biochemistry, which can have subsequent consequences on protein structure, generating a loss-of-function mutation. However, a second-site correcting mutation can permit a reversion of the altered protein's biochemistry to its original nature, most likely by restoring local structure and thereby restoring protein function. Amino acid changes in a protein generally induce local structural changes. Therefore, it seems likely that second site suppressors or correcting mutations will lie close in the protein's three-dimensional native structure to the original mutation.

Studies dating back to the 1990s have documented occasional second-site amino acid substitutions in residues proximal to the originally substituted residue (Chen, Dean, Grobler, & Hurley, 1996; Hartzog & Cain, 1994; Perwez et al., 2008; Shen, Sosa-

Peinado, & Mueller, 1996). Recently, the scientific community has realized that this complementing phenomenon can be exploited to provide valuable assistance when predicting the structures of proteins from their amino acid sequences. Current work has been able to use only sequencing data in conjunction with the Rosetta protein folding program to quite accurately predict the three-dimensional structures of folded peptides (Bitbol, Dwyer, Colwell, & Wingreen, 2016; Ovchinnikov et al., 2017). These findings provide confidence that complementing amino acid changes often lie near each other, if not directly interact. The identification of complementing substitutions in AraC could provide information about inter-domain interactions. In turn, these interactions could explain the protein's conformational changes upon the binding of arabinose and how the changes result in the induction or repression of transcription. This approach requires a means of mutating residues and quickly assessing protein function.

Materials and Methods (Chapter 4)

Plasmids

The pBAD-GFP (*amp*^{r}) plasmid was previously described as pBAD-AraC-GFP (Timmes et al., 2004). The plasmid was briefly sold by Clontech® (Mountain View, CA). It contains the *GFP* gene driven by the P_{BAD} promoter, the *araC* gene driven by the P_C promoter, and the ampicillin resistance gene β-lactamase.

The pCCN-GFP and pCCN-mCherry (*kan*^{r}) plasmids were derived from iGEM pSB2K4 (Cambridge, MA) and named CCN for their controllable copy number. The plasmids contain the full P_{BAD} regulatory region from pBAD-GFP, including the *araC* gene under the P_C promoter and *GFP* or *mCherry* under the P_{BAD} promoter (Rodgers,

unpublished). pCCN-GFP and pCCN-mCherry can be maintained at single copy number with elevated levels of *lac* repressor (LacI). Alternatively, they can be increased to high copy number for sequencing if cells containing pCCN are induced with Isopropyl β-D-1-thiogalactopyranoside (IPTG).

The pSE380 (*amp*^{r}) plasmid was obtained from Invitrogen (San Diego, CA). pSE380 is a high copy number plasmid containing the *lacI* gene, used in this study to keep pCCN-GFP and pCCN-mCherry at presumably single copy number.

Strains

SH321 (*F*$^-$*, ΔaraC-leu1022 Δlac74, galK*$^-$*, str*^{r}) cells lack a chromosomal copy of AraC, as described previously (Hahn et al., 1984).

RS321 (*SH321 by P1 transduction, with TN10-I*^{Q}*-lac, cm*^{r}*::lac*) cells contain the *lacI* gene regulated by the *lacI*^{Q} promoter, used for controlling the copy number of pCCN-GFP and pCCN-mCherry (Schleif, unpublished).

Competent XL1-Red® cells were obtained from Agilent Technologies (Santa Clara, CA). These cells contain mutator alleles and have inactivation in three DNA repair pathways. As a result, they generate insertions, deletions, and base changes with a frequency approximately 5000-fold higher than typical *E. coli* cells (Reed & Schleif, 1999).

Media

Yeast tryptone (YT) was used for rich medium as previously described, both for plates and liquid cultures. Plates were made with 15 grams of agar per liter. (R. F. Schleif

& Wensink, 1981). To screen cells using a single sugar source, minimal media were used: either A+B or M10, as previously described (R. F. Schleif & Wensink, 1981). Minimal media were supplemented with leucine (50 μg/mL), vitamin B1 (10 μg/mL), $MgSO_4$ (1 mM), and either arabinose or glycerol (2 mg/mL). IPTG was used at a concentration of 1 mM unless otherwise indicated. Antibiotics were used at the following concentrations unless otherwise indicated: ampicillin 100 μg/mL (amp), kanamycin 40 μg/mL (kan).

Cell Growth

Cells were grown at 37 ˚C unless otherwise indicated. For liquid media, cells were grown in a shaker to approximately stationary phase unless otherwise indicated. Typical growth duration was 1 day on YT medium and 2-4 days on minimal media.

Site-Directed Mutagenesis

The Agilent Technologies QuikChange™ protocol for oligonucleotide-directed mutagenesis was followed for all reactions. After thermocycling, the reaction mixture was incubated with *Dpn*I (1 μL, 37 ˚C, 2 hours) obtained from New England Biolabs® (Ipswich, MA) to digest any parental (methylated) DNA. Primers were ordered from Integrated DNA Technologies (Coralville, IA).

Transformation

Plasmid DNA (1-2 μL) was mixed with 50 μL competent *E. coli* cells (various strains) and incubated on ice for 30 min. Cells were subjected to heat-shock (42 ˚C, 20-45

seconds) and promptly returned to ice (2 min). YT medium (500 µL) was added, and cells were allowed to recover without antibiotics (37 ˚C, approx. 1 hour). Following recovery, cells were either sterilely spread onto appropriate plates or used to inoculate liquid media (YT unless otherwise indicated).

Plasmid DNA Extraction

Plasmid DNA was extracted from cells using the Promega Wizard® *Plus* SV Minipreps DNA Purification System: Cells were used to inoculate 5 mL liquid YT media supplemented with appropriate antibiotics and other nutrients if necessary. Tubes were incubated until the cells reached sufficient growth for sequencing (37 ˚C, approx. 18 hours). Cells were spun down (3000 RPM, 10 min), and plasmid DNA was extracted following the protocol. DNA concentration was determined using a NanoDrop 2000 UV-Vis Spectrophotometer.

Sequencing

Plasmid DNA (approx. 650 ng) was sent to GENEWIZ® (South Plainfield, NJ) for sequencing with the appropriate primer.

Developing a System to Screen for Mutants

The plasmid pBAD-GFP contains the gene encoding GFP driven by the P_{BAD} promoter, the *araC* gene driven by the P_C promoter, and the ampicillin resistance gene β-lactamase (Figure 3). The *E. coli* strain SH321 lacks a chromosomal copy of *araC*. Transforming pBAD-GFP into SH321 cells creates a system in which *araC* (on the

plasmid) can be easily mutagenized and sequenced, while the growth of cells and protein function can be simultaneously observed *in vivo*. Because AraC induces the P_{BAD} promoter in the presence of arabinose, AraC function should result in proportional induction of *GFP* from P_{BAD}. Thus, GFP fluorescence can be used as an indication of AraC functionality.

In addition to GFP, this system provides another means of assessing AraC function: cell growth. Functional AraC is necessary for *E. coli* cells to metabolize arabinose, since it is the regulatory protein of the operon and is necessary to induce the transcription of *araBAD*. If cells are grown in minimal media containing arabinose as the sole sugar source, then AraC function should also be proportional to cell growth – in addition to *GFP* induction – because AraC is needed to metabolize arabinose. The better that cells can metabolize arabinose, the better they can grow. The dual role of the plasmid-based AraC provides two methods of assaying activity: cell growth and whole-cell GFP fluorescence. Notably, both growth and GFP fluorescence are qualitative measures of activity unless a fluorescence assay is performed. Cell growth and fluorescence can be used to indicate whether AraC is relatively functional or non-functional, but should not be used quantitatively. For the purposes of identifying loss-of-function mutants and corrected mutants, however, a heuristic measure is sufficiently informative. In fact, the relative ease of assessment makes a heuristic measure advantageous.

Generating Primary Mutations in *araC*

pBAD-GFP was mutagenized by transformation into the mutating *E. coli* strain XL1-Red®, which has a random mutation rate approximately 5000-fold higher than common strains due to its deficiency in DNA repair pathways. Cells were plated on YT media containing 100 μg/mL ampicillin (YT/amp) to select for transformants. Resulting colonies were scraped off the plate, and their plasmid DNA was extracted using the Promega Miniprep protocol. Extracted plasmids had random mutations throughout due to the error-prone DNA polymerase of XL1-Red®. This heterogeneous mixture of plasmids was expected to contain some plasmids with loss-of-function mutations in *araC*.

Identifying Cells with Loss of Function Mutations

Mutated plasmids were transformed into the *araC*$^-$ cell line SH321; cells were grown on YT/amp at 37 ˚C overnight. As a control, wild-type pBAD-GFP was also transformed into SH321 cells and grown on YT/amp overnight ("WT cells/colonies"). After growth of the mutant colonies, replica plating was performed onto a plate of minimal media containing arabinose (2 mg/mL) as the sole sugar source. These plates also contained M10 salts, leucine (50 μg/mL), vitamin B1 (10 μg/mL), $MgSO_4$ (1 mM), and ampicillin (100 μg/mL) – abbreviated as min/ara/amp. After 2 days of growth at 37 ˚C, colonies on the replica plate were assessed for growth and GFP fluorescence. Colonies with decreased growth compared to others were streaked onto a fresh plate of YT/amp (incubated at 37 ˚C overnight) to ensure isolation of a single colony.

Isolated colonies were dual-spotted onto two types of media: min/ara/amp and min/gly/amp (same as min/ara/amp, but with glycerol instead of arabinose), in that order.

The order is important because the first spot may get more cells than the second. To avoid mistakenly identifying a colony as having poor growth on arabinose, it is important to spot onto arabinose first. WT colonies were also spotted as a control. After incubation for 2 days at 37 ˚C, colony growth and fluorescence was observed; those with decreased growth on arabinose ("AraC$^-$") compared to WT colonies were selected for subsequent steps. Fluorescence was used merely as an aid in identifying poorly-growing colonies, since a non-fluorescent colony could have a loss-of-function mutation in *GFP* instead of *araC*. Cell growth was used as the final indication of AraC functionality.

Generating Second-Site Correcting Mutations

Sterile tubes of 5 mL liquid min/gly/amp were individually inoculated with AraC$^-$ colonies and grown at 37 ˚C. During this growth, mutations were randomly generated throughout the pBAD-GFP plasmids due to the normal error of *E. coli* DNA polymerase. After two days, cells were spun down (3000 RPM, 10 min), resuspended in 100 µL min/gly/amp, plated onto min/ara/amp, and grown for approximately 2 days at 37 ˚C. Colonies with prominent growth were streaked on YT/amp and incubated at 37 ˚C overnight. Isolated colonies were dual-spotted onto min/gly/amp and min/ara/amp, in that order. WT colonies were also spotted as a control. After incubation for 2 days at 37 ˚C, colonies with WT phenotype ("AraC$^+$" revertants) were selected.

Sequencing AraC$^+$ Revertants

AraC$^+$ revertant colonies were used to individually inoculate sterile tubes of 5 mL YT/amp; tubes were incubated at 37 ˚C overnight. After growth, cells were spun down

(3000 RPM, 10 min), and plasmid DNA was extracted using the Promega Miniprep protocol. The plasmid was sequenced to obtain the AraC coding sequence. Sequences were analyzed for the presence of a primary and correcting mutation. Figure 4 outlines the entire process for generating AraC$^-$ mutations and correcting mutations.

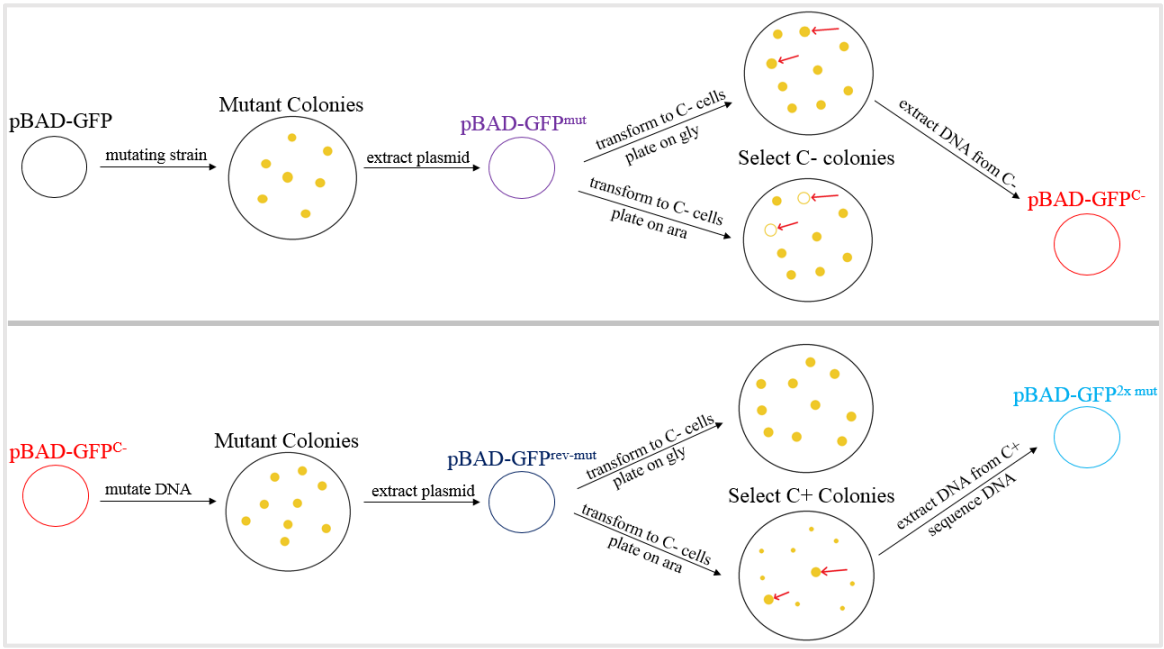

Figure 4. Chapter 4 methods overview. pBAD-GFP DNA was passed through the mutating strain XL1-Red® to generate loss-of-function mutations in *araC*. Cells with mutations were selected using differential growth on minimal ara versus minimal gly media. Plasmid DNA was extracted from cells with poor or no growth on ara. Second-site correcting mutations were generated during subsequent cell growth. Colonies were spotted on minimal ara and minimal gly once more; plasmid DNA was extracted from cells with WT-like growth on ara. The extracted DNA was then sequenced and analyzed for the presence of complementing mutations.

Results and Discussion (Chapter 4)

The method described generated mutations in a*raC* on the pBAD-GFP plasmid. Sequencing results verified whether mutagenesis screens were successful in generating primary and secondary substitutions. The above procedure was repeated three times (mutagenesis screens 1-3); however, revertants were not generated in screen 3 (discussed below). In mutagenesis screen 4, a new plasmid (pCCN) was used.

Mutagenesis Screen 1

Five AraC$^+$ revertant colonies were sequenced (Table 5). The screen generated one colony with two amino acid substitutions in AraC, two colonies with a single amino acid substitution in AraC, and two colonies with WT AraC.

Table 5. AraC$^+$ revertants that were sequenced in mutagenesis screen 1.

Colony	Mutations	Amino Acid Changes
1	81(CAC → CAA)	H81Q
	146(CGC → CCC)	R146P
2	12(GGA → AGA)	G12R
3	156(CTT → GTT)	L156V
4	Wild Type	
5	Wild Type	

H81Q and R146P

Histidine 81 lies within the β-sheet of AraC's dimerization domain (DD). Previous studies have shown that in the presence of arabinose, the AraC mutant H81P induces P_{BAD} at only 20% of the induction seen in WT cells. On the other hand, H81P AraC represses P_C, the promoter for *araC*, 100% compared to WT (Wendy Reed, unpublished). Because AraC simultaneously controls the transcription off P_C and P_{BAD}, repression of P_C is expected to be reflected by repression of P_{BAD} as well. Arginine 146 lies in the loop between two α-helices of the DD. The R146P mutant has not been documented. However, as discussed in Chapter 1, many mutations in the loop can be tolerated without affecting AraC function.

Given that mutation of H81 has been shown to reduce induction, it is likely that H81Q was the original mutation to produce an AraC$^-$ phenotype. R146P could be a

second-site correcting mutation, or it could be a trivial mutation that occurred at any point, given that the loop has been mutated without functional consequences. Residues 81 and 146 are both in the dimerization domain, which has a rigid structure that has been crystallized. These residues are not proximal to each other, so it is unlikely that they would interact directly. It is possible that far-away residues could also complement each other, but a more likely explanation is that the R146P mutation is insignificant to AraC's structure and function. Decreased induction due to H81Q could still be sufficient to produce WT-like growth with the high copy number pBAD-GFP plasmid, allowing the double mutant H81Q/R146P to pass as a pair of complementing mutations. Introducing the R146P substitution alone into AraC could verify that the substitution is functionally insignificant.

G12R

G12 lies on the N-terminal arm of AraC. Previous studies have determined that the G12R mutant is constitutive, but inducible (Jennifer Ross, Thesis). In the absence of arabinose, G12R induces transcription from the P_{BAD} promoter 100% relative to the amount of induction from WT protein in the presence of arabinose. G12R further induces transcription to 200% of WT in the presence of arabinose. Without a second mutation, this colony is not useful for the study.

L156V

Residue 156 lies in an α-helix of the DD. Previous studies have determined that the L156V mutant is low-level constitutive, meaning that it induces transcription even in

the absence of arabinose (Schleif, unpublished). Again, without a second mutation, this colony is not useful for the study.

WT

Two colonies were WT in sequence, indicating either that the selection process was imperfect, or the original mutations were directly reverted back to WT sequence. In mutagenesis screen 1, loose criteria were used to select for $AraC^-$ cells; cells with reduced growth on min/ara were still considered, as opposed to only considering cells with absolutely no growth. WT cells could have passed through the screening process for $AraC^-$ candidates. Alternatively, the original mutation that produced the $AraC^-$ phenotype could have reverted back to wild type.

Mutagenesis Screen 2: Stricter $AraC^-$ Selection

In mutagenesis screen 2, I used more rigid $AraC^-$ selection criteria. Only cells that had no growth on arabinose were pursued for generating $AraC^+$ revertants. Five more $AraC^+$ revertants were sequenced (Table 6). The use of strict select criteria for $AraC^-$ colonies likely caused nonsense or frameshift mutations to be selected for, rather than single-amino acid substitutions. Substitutions are less likely to cause severe changes in protein function. The problem of selecting for more frameshift and nonsense mutations is exacerbated with a high copy number plasmid. At high copy number, a significant but not complete loss of AraC function from a substitution can be outweighed by having more protein present. Therefore, a loss-of-function mutant can appear to be phenotypically

wild type. Additionally, four colonies in this screen possessed the same genotype; this is probably due to the selection of colonies descended from the same transformant.

Table 6. AraC$^+$ revertants that were sequenced in mutagenesis screen 2.

Colony	Mutations	Amino Acid Changes
6	84(CGT → CAT) 104(TGG → TGA)	R84H W104STOP
7	84(CGT → CAT) 104(TGG → TGA)	R84H W104STOP
8	84(CGT → CAT) 104(TGG → TGA)	R84H W104STOP
9	84(CGT → CAT) 104(TGG → TGA)	R84H W104STOP
10	62(CGA → CGT) 54(CAG → CA–)	R62R Q54frameshift

R84H and W104STOP

Four colonies had the same sequence; two were a pair of revertants from one AraC$^-$ strain, and the other two were a pair of revertants from a different AraC$^-$ strain. Because there are limited options for residues close enough to complement an AraC$^-$ mutation, it is possible that two revertant colonies with the same AraC$^-$ mutation would also have the same correcting mutation. However, it is unlikely that 4 revertants would all share genotypes if they already had unique AraC$^-$ mutations before secondary mutations were generated. The repetition seen here implies that the AraC$^-$ colonies selected were descendants of the same transformant when mutated plasmid was transformed into SH321 cells in the AraC$^-$ generation steps. To minimize duplication of transformants, I decided to separate colonies immediately after the heat-shock step in future transformations and only select a single AraC$^-$ cell from each separation.

The most likely explanation for the genotypes seen in these revertants is that the primary mutation was W104STOP, producing a completely non-functional AraC and leading to no growth on arabinose. The correcting mutation could be a nonsense-suppressor on the *E. coli* chromosome – which would not appear in plasmid sequencing. Finally, R84H could have an insignificant effect on AraC function. R84H has not previously been characterized, but both arginine and histidine residues are basic, and most single substitutions do not significantly affect protein function.

R62R, Q54frameshift

This colony had a silent mutation and a frameshift mutation. The Q54frameshift was likely the primary mutation, and it would be expected to produce a non-functional AraC. A secondary correcting mutation could be a frameshift-suppressing mutation on the *E. coli* chromosome. Finally, the R62R mutation is silent and thus would not be predicted to affect AraC function.

The goal of this approach was to generate single-amino acid substitutions to study the structure and function of AraC. Frameshifts and nonsense mutations likely appeared in mutagenesis screen 2 due to the stricter selection criteria for $AraC^-$ colonies. A single-amino acid change is unlikely to completely knock down protein function in most proteins, including AraC. Thus, using strict selection criteria for $AraC^-$ colonies seems unlikely to produce the desired type of mutations – at least when using a high copy number plasmid.

Mutagenesis Screen 3: Lenient AraC$^-$ Selection

In mutagenesis screen 3, the AraC$^-$ colonies were sequenced instead of the AraC$^+$ revertants to verify that the AraC$^-$ generation process was working before attempting to identify revertants. AraC$^-$ mutations needed to be single amino acid substitutions so that revertants could possess two complementing substitutions. For this screen, the AraC$^-$ selection criteria was relaxed once more. A total of six AraC$^-$ candidates were sequenced without progressing to the AraC$^+$ revertant stage (Table 7). The problem of related descendants persisted in the screen, indicating that the issue involves the AraC$^-$ selection step. Remarkably, one colony was found to have a WT AraC sequence despite having reduced growth on min/ara/amp. On the other hand, two other colonies were able to partially grow on min/ara/amp despite having a frameshift mutation at residue 97; they are discussed below.

Table 7. AraC$^-$ candidates that were sequenced in mutagenesis screen 3.

Colony	Mutations	Amino Acid Changes
11	97(TAC → TA–)	Y97frameshift
12	97(TAC → TA–)	Y97frameshift
13	Wild Type	
14	83(GGT → GCT) 214(CTT → TTT) 290(AAG → GAG)	G83A L214F K290E
15	83(GGT → GCT) 214(CTT → TTT) 290(AAG → GAG)	G83A L214F K290E
16	83(GGT → GCT) 214(CTT → TTT) 290(AAG → GAG)	G83A L214F K290E

Y97frameshift

This mutation is interesting because a frameshift so early in the protein sequence should render a completely non-functional AraC. However, both colonies exhibited some growth on arabinose, although less than WT. Therefore, I postulate that some other mutation occurred in addition to the frameshift. One explanation would be an in-frame translation restart after the frameshift, producing two polypeptide fragments that could later combine into a functional AraC (Eustance & Schleif, 1996). However, there are neither possible restart codons after the frameshift nor STOP codons near residue 97 either with or without the frameshift. With this restart model excluded, further study would be needed to understand colonies possessing Y97frameshift. The most likely explanation is that there is some complementing phenomenon occurring, but not visible in the coding sequence of the *araC* on the plasmid. There could be a mutation elsewhere on the plasmid or possibly on the *E. coli* chromosome.

WT

One colony demonstrates the risk of using lenient AraC$^-$ selection – a WT colony could be selected as AraC$^-$ when looking for marginally decreased growth. Alternatively, there may be mutations outside of the AraC coding sequence that contribute to decreased growth – such as a mutation in the P_C promoter or a mutation on the *E. coli* chromosome.

G83A, L214F, and K290E

None of the three mutations found has been previously characterized. The triple mutation is not directly useful for this study because it is impossible to know which of the

three mutations contributed to decreased AraC function. This triple mutant, however, may be worthwhile to pursue in future studies. Before progressing to the stage of generating correcting mutations, the process for creating $AraC^-$ mutations must be optimized because none of the 6 $AraC^-$ candidates sequenced in mutagenesis screen 3 had a single amino acid substitution.

Using a High Copy Number Plasmid Introduces Confounding Data

The copy number of pBAD-GFP is several hundred plasmids per cell. A decrease in AraC function may not result in a proportional decrease in P_{BAD} induction – and thus growth – if *araC* is expressed from many plasmids. A sufficient amount of partially active AraC could provide enough induction for a WT growth phenotype. Additionally, having multiple plasmids poses a problem if they are not identical. In the method used, $AraC^+$ revertants were generated through random mutation in cells that contained hundreds of pBAD-GFP plasmids. Each copy would mutate independently, so the extracted plasmid DNA could be heterogeneous. Transformation and using a low copy number plasmid are two approaches to mitigate the problem. During transformation, bacterial cells are generally expected to take up a single plasmid, particularly if standard protocols are employed.

The mechanism of transformation ensures that only a single plasmid will be present in a cell until replication is allowed to occur. Without any selective pressures, only a trivial number of plasmids extracted and sequenced after growth should have additional mutations. However, a more comprehensive solution would be to use a low copy number plasmid.

Mutagenesis Screen 4: Using a Plasmid with Controllable Copy Number

The pCCN plasmids (pCCN-GFP and pCCN-mCherry) contain the P_{BAD} regulatory region, but they can be maintained at controllable copy number in the presence of the *lac* repressor, supplied by pSE380. To perform screens using pCCN, pSE380 was transformed into competent SH321 cells. pCCN was then transformed into the resulting SH321-pSE380 cells, and primary amino acid substitutions were generated as described in the method above, without progressing to generating revertants. Before sequencing, plasmid DNA was digested with restriction enzymes unique to pSE380. The remaining pCCN was re-transformed into SH321 cells lacking pSE380, then cells were grown in 5 mL YT/kan. Plasmid DNA was extracted, and the *araC* genes were sequenced. Seven $AraC^-$ colonies were sequenced (Table 8).

Table 8. $AraC^-$ candidates that were sequenced in mutagenesis screen 4.

Colony	Mutations	Amino Acid Changes
17	66(TGC → TAC)	C66Y
18	114(TTT → CTT)	F114L
19	114(TTT → CTT)	F114L
20	114(TTT → CTT)	F114L
21	114(TTT → CTT) 229(GAC → AAC)	F114L D229N
22	33(TTT → TTTT)	F33frameshift
23	270(GGG → GG–)	G270frameshift

C66Y

Residue 66 lies buried in the β-barrel of the dimerization domain. Due to the cysteine originally at position 66, this residue could play a role in protein folding. It is

plausible that removal of the cysteine may result in an improperly folded protein and thus an AraC$^-$ phenotype. This mutation is a possible candidate for reversion in future studies.

F114L

Residue 114 lies near the β-barrel of the dimerization domain, but is partially exposed. The side chain of residue 114 may interact with other residues within the β-barrel. This residue is a good candidate for reversion in future studies.

F114L and D229N

F114L was described above; the single amino substitution is sufficient to produce an AraC$^-$ phenotype. Residue 229 lies in the DNA-binding domain of AraC and is not well characterized. Without further study, it is not possible to say whether F114L and D229N act together to produce an AraC$^-$ phenotype, whether D229N causes an AraC$^-$ phenotype on its own, or whether D229N produces no functional change, and the phenotype observed was only due to F114L.

F33frameshift

This frameshift involved the introduction of an additional thymine in the 7-thymine sequence corresponding to residues 33-35. This mutation would be expected to produce an AraC protein that has the correct amino acid sequence for residues 1-35, an incorrect amino acid sequence for residues 36-69, and termination after residue 69 due to the frameshift introducing and early STOP codon.

G270frameshift

This frameshift was due to the deletion of a guanine at residue 270. This mutant likely loses its ability to bind DNA, as the final α-helix of the DNA-binding domain is disrupted. The mutation present here would be expected to produce the correct amino acid sequence for residues 1-270, an incorrect sequence for residues 270-283, and termination after residue 283.

Refining of the Selection System

In mutagenesis screen 4, two frameshift mutations occurred in regions with repeated nucleotides (*e.g.* 7-thymine sequence). Frameshifts likely occur at high frequencies in these regions due to increased DNA polymerase slippage. Even with a single copy number plasmid, frameshifts may persist in the generation of $AraC^-$ mutations. However, the use of pCCN successfully generated two mutants with single amino acid substitutions. A controllable copy number plasmid appears to be the best construct for this genetic method.

To further refine the screening system by eliminating the high copy number pSE380 plasmid, the cell line RS321 was developed (Schleif, unpublished). RS321 cells contain the *lacI* gene and I^Q promoter, allowing for control of pCCN copy number using only IPTG. Preliminary tests in the system demonstrated that RS321 functions as expected. RS321 cells containing pCCN-GFP and pCCN-mCherry grown with arabinose and IPTG produced significant fluorescence from GFP or mCherry. Cells grown with only arabinose, only IPTG, or neither did not produce visible fluorescence. Future studies should continue the generation of $AraC^-$ colonies and $AraC^+$ revertants in this system.

Summary of Mutagenesis Screens (Chapter 4)

My proposed method of generating complementing amino acid substitutions in AraC establishes a foundation upon which future studies may build. The system of using a controllable copy number plasmid to encode *araC* and the *lacI*^{Q} strain RS321 to conduct genetic screens is one that has been developed with significant thought after many experiments in other systems. Although pairs of complementing substitutions have not yet been consistently identified in AraC, the success of this genetic approach in studying other proteins suggests that it can be applied to AraC. The complementing substitutions method holds promise for answering questions about AraC where other approaches have failed. Once my method is completely refined, the payoff – both to research of AraC and of other proteins – will be considerable.

REFERENCES

Bitbol, A., Dwyer, R. S., Colwell, L. J., & Wingreen, N. S. (2016). Inferring interaction partners from protein sequences. *Proceedings of the National Academy of Sciences, 113*(43), 12180-12185. doi:10.1073/pnas.1606762113

Chen, R., Dean, A. M., Grobler, J. A., & Hurley, J. H. (1996). Second-site suppression of regulatory phosphorylation in escherichia coli isocitrate dehydrogenase. *Protein Science, 5*(2), 287-295. doi:10.1002/pro.5560050213

Eustance, R. J., & Schleif, R. F. (1996). I*n vivo* association of protein fragments giving active AraC. *Proteins, 25*(4), 501-505. doi:10.1002/prot.9

Hahn, S., Dunn, T., & Schleif, R. (1984). Upstream repression and CRP stimulation of the *escherichia coli* L-arabinose operon. *Journal of Molecular Biology, 180*(1), 61-72.

Hahn, S., & Schleif, R. (1983). *In vivo* regulation of the *escherichia coli araC* promoter. *Journal of Bacteriology, 155*(2), 593-600.

Harmer, T., Wu, M., & Schleif, R. (2001). The role of rigidity in DNA looping-unlooping by AraC. *Proceedings of the National Academy of Sciences of the United States of America, 98*(2), 427-431. doi:10.1073/pnas.98.2.427

Hartzog, P. E., & Cain, B. D. (1994). Second-site suppressor mutations at glycine 218 and histidine 245 in the alpha subunit of F1F0 ATP synthase in escherichia coli. *Journal of Biological Chemistry, 269*(51), 32313-32317.

Khlebnikov, A., Risa, Ø, Skaug, T., Carrier, T. A., & Keasling, J. D. (2000). Regulatable arabinose-inducible gene expression system with consistent control in all cells of a culture. *Journal of Bacteriology, 182*(24), 7029-7034. doi:10.1128/JB.182.24.7029-7034.2000

Lobell, R. B., & Schleif, R. F. (1990). DNA looping and unlooping by AraC protein. *Science (New York, N.Y.), 250*(4980), 528-532.

Malaga, F., Mayberry, O., Park, D., Rodgers, M. E., Toptygin, D., & Schleif, R. F. (2016). A genetic and physical study of the interdomain linker of E. coli AraC protein-a trans-subunit communication pathway. *Proteins: Structure, Function, and Bioinformatics, 84*, 448--460.

Ovchinnikov, S., Park, H., Varghese, N., Huang, P., Pavlopoulos, G. A., Kim, D. E., . . . Baker, D. (2017). *Protein structure determination using metagenome sequence data* doi:10.1126/science.aah4043

Perwez, T., Hami, D., Maples, V. F., Min, Z., Wang, B., & Kushner, S. R. (2008). Intragenic suppressors of temperature-sensitive rne mutations lead to the dissociation of RNase E activity on mRNA and tRNA substrates in escherichia coli. *Nucleic Acids Research, 36*(16), 5306-5318. doi:10.1093/nar/gkn476

Reed, W. L., & Schleif, R. F. (1999). Hemiplegic mutations in AraC protein. *Journal of Molecular Biology, 294*(2), 417-425. doi:10.1006/jmbi.1999.3224

Rodgers, M. E., Holder, N. D., Dirla, S., & Schleif, R. (2009). Functional modes of the regulatory arm of AraC. *Proteins, 74*(1), 81-91. doi:10.1002/prot.22137

Rodgers, M. E., & Schleif, R. (2009). Solution structure of the DNA binding domain of AraC protein. *Proteins, 77*(1), 202-208. doi:10.1002/prot.22431

Saviola, B., Seabold, R., & Schleif, R. F. (1998). Arm-domain interactions in AraC. *Journal of Molecular Biology, 278*(3), 539-548. doi:10.1006/jmbi.1998.1712

Schleif, R. (2003). AraC protein: A love-hate relationship. *BioEssays : News and Reviews in Molecular, Cellular and Developmental Biology, 25*(3), 274-282. doi:10.1002/bies.10237

Schleif, R. (2010). AraC protein, regulation of the L-arabinose operon in *escherichia coli*, and the light switch mechanism of AraC action. *FEMS Microbiology Reviews, 34*(5), 779-796. doi:10.1111/j.1574-6976.2010.00226.x

Schleif, R. F., & Wensink, P. C. (1981). *Practical methods in molecular biology*. New York: Springer-Verlag.

Shen, H., Sosa-Peinado, A., & Mueller, D. M. (1996). Intragenic suppressors of P-loop mutations in the -subunit of the mitochondrial ATPase in the yeast saccharomyces cerevisiae. *Journal of Biological Chemistry, 271*(20), 11844-11851. doi:10.1074/jbc.271.20.11844

Soisson, S. M., MacDougall-Shackleton, B., Schleif, R., & Wolberger, C. (1997). Structural basis for ligand-regulated oligomerization of AraC. *Science (New York, N.Y.), 276*(5311), 421-425.

Studier, F. W., Rosenberg, A. H., Dunn, J. J., & Dubendorff, J. W. (1990). Use of T7 RNA polymerase to direct expression of cloned genes. *Methods in Enzymology, 185*(Gene Expression Technol.), 60-89.

Timmes, A., Rodgers, M., & Schleif, R. (2004). Biochemical and physiological properties of the DNA binding domain of AraC protein. *Journal of Molecular Biology, 340*(4), 731-738. doi:10.1016/j.jmb.2004.05.018

Weldon, J. E., Rodgers, M. E., Larkin, C., & Schleif, R. F. (2007). Structure and properties of a truely apo form of AraC dimerization domain. *Proteins, 66*(3), 646-654. doi:10.1002/prot.21267

Wu, M., & Schleif, R. (2001a). Mapping arm-DNA-binding domain interactions in AraC. *Journal of Molecular Biology, 307*(4), 1001-1009.

Wu, M., & Schleif, R. (2001b). Strengthened arm-dimerization domain interactions in AraC. *The Journal of Biological Chemistry, 276*(4), 2562-2564. doi:10.1074/jbc.M008705200

CURRICULUM VITAE

ARJUN TAMBE

Born on March 31, 1995 in Barrington, IL.

EDUCATION

University of Illinois, Chicago, IL: Aug 2017 – May 2021
Pursing an M.D.

Johns Hopkins University, Baltimore, MD: Aug 2016 – May 2017
M.S. in Molecular Biology; includes coursework, research, teaching, and a master's thesis.

Johns Hopkins University, Baltimore, MD: Aug 2013 – May 2016
B.S. in Molecular & Cellular Biology, early graduation, minor in Music.
Graduated with general honors and departmental honors.

Illinois Mathematics and Science Academy (IMSA), Aurora, IL: Aug 2010 – June 2013

RESEARCH EXPERIENCE

Johns Hopkins University – Homewood Campus: May 2014 – Present
Schleif lab, Department of Biology.

Research on the mechanisms of AraC and the regulation of the L-arabinose operon in *E. coli*. Data found that mutation in the loop region of AraC did not significantly affect ability to induce or repress gene expression. Developed a fluorescence-based assay for long-term lab use, providing a means of quickly measuring gene induction without the use of enzyme assays. Also developed a method of identifying intra- and inter-domain interactions through the use of complementing mutations.

Presented results at Hopkins' undergraduate research day.

University of Chicago, Chicago, IL – Gordon Center for Integrative Science: Aug 2011 – May 2013
Jean T. Greenberg lab, Department of Molecular Genetics and Cell Biology.

Studied the effects of azelaic acid on root growth and gene expression in *Arabidopsis thaliana*. Data identified two genes, *EARLI1* and *SFD1*, and suggested their implication in plants' natural defense pathways in response to pathogenic infection.

Presented results at IMSAloquium, NCSSSMST, and JSHS research conferences; presented at the Illinois Junior Academy of Science (IJAS) regional and state competitions, received the IJAS gold award. Submitted for publication in MPMI.

RELATED EXPERIENCE

Teaching Assistant – Johns Hopkins University: Jan 2015 – Present
Conduct weekly sections for statistics and biology courses; help students outside of class; grade assignments exams. Students who attend the sections perform significantly better on homework and exams. On mandatory teaching assistant evaluations, the students' rating averages 4.65 out of 5.

Private Tutoring: May 2013 – Present
Tutor various subjects including SAT/ACT preparation, natural sciences, and college essay writing. Students' performance always improves after tutoring. The average rating of lessons is 4.95 out of 5.

Shadowing in Nephrology: Aug 2015 – Dec 2015
Shadowed Bernard Jaar, a JHU doctor of nephrology, weekly at Good Samaritan Hospital in Baltimore, MD.

Shadowing in Internal Medicine: July 2013 – Aug 2013
Shadowed Rathna Kumar Yallapragada, a doctor of internal medicine in Hinsdale, IL.

LEADERSHIP EXPERIENCE

The JHU Mental Notes (*a cappella* Group) – Director: Aug 2013 – Present
Elected director in sophomore year and was unanimously re-elected the following year. Responsibilities include leading rehearsals 4 times per week; arranging and teaching music. Supervise group conduct, including creating rehearsal polices and outlining expectations. Perform every 1-2 weeks.

JHU Performing Arts Committee (PAC) Co-Chair: May 2014 – Present
Manage the Hopkins performing arts groups. Coordinate rehearsal rooms; facilitate inter-group relations and mediate conflicts; assist the JHU Director of Arts in planning large events such as annual shows featuring all arts groups.

JHU Cricket Club – Co-founder/President: Sep 2014 – Present
Founded a cricket club for Hopkins students to practice and view matches regularly. Began leading the club as president after formal approval of the organization; the club now has 40+ members. Hold events to spread knowledge of the game and teach others how to play.

VOLUNTEER ACTIVITIES

Blue Key Society – Johns Hopkins University (Tour Guide): Dec 2013 – Mar 2016
Give campus admissions tours and answer questions about the university. Tour groups range in size, sometimes exceeding 50. Also host prospective and admitted students overnight, educating them from a personal perspective.

Boy Scouts of America – Pack 153, Troop 335: Sep 2000 – Aug 2013
Held various troop positions such as troop guide, webmaster, and patrol leader. Advanced through the ranks of scouting to Eagle Scout, the highest award. The Eagle Project involved fundraising $1900 for materials and miscellaneous costs, then constructing improvements to a local school over the course of two weekends.

Ethics Facilitator: May 2012 – May 2013
Taught a class of 25 students about various ethical systems in monthly sessions over the course of the school year. Stressed the importance of living a life of virtue and making ethical decisions without considering personal gain.

AWARDS

Eagle Scout, BSA Troop 335	2012
US Presidential Scholarship 2013 Candidate	2013
Illinois Junior Academy of Science State Exposition – Gold Award	2013
National Merit Scholarship Finalist	2013
AP Scholar with Distinction	2013
Illinois State Scholar	2013